DOCUMENT RESUME

ED 358 133 TM 019 916

AUTHOR Centra, John A.
TITLE Use of the Teaching Portfolio and Student Evaluations for Summative Evaluation.
PUB DATE Apr 93
NOTE 34p.; Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *College Faculty; College Students; Community Colleges; Correlation; Deans; Educational Cooperation; *Faculty Evaluation; *Portfolios (Background Materials); *Student Evaluation of Teacher Performance; *Summative Evaluation; Supervisors; Teacher Characteristics; *Teaching Skills; Two Year Colleges
IDENTIFIERS *Performance Based Evaluation

ABSTRACT

The use of the teaching portfolio and student evaluations in evaluating 97 faculty members at a community college for contract renewal was studied. Two faculty peers and a dean evaluated the portfolios of each teacher. Deans also visited classrooms. Portfolios could include material about students that reflected their learning, material from the faculty member, and material from others with a bearing on teacher performance. The student evaluations correlated reasonably well and on similar teaching dimensions with evaluations by the deans and one of the peers. Recommendations for the construction and evaluation of teaching portfolios as well as for the use of student evaluations in summative evaluations are offered. The faculty in this study did not have the opportunity to put portfolios together over several years, but even portfolios that were not ideally designed assisted evaluators of teaching performance. Six tables present study findings. (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED358133

Use of the Teaching Portfolio and Student Evaluations for Summative Evaluation

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

John A. Centra
Professor and Chair
Program of Higher Education
Syracuse University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
JOHN A. CENTRA
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Abstract

This study investigated the use of the teaching portfolio and student evaluations in evaluating 97 faculty members for contract renewal. Two faculty peers and a dean evaluated the portfolios of each teacher; deans also visited classrooms. The student evaluations correlated reasonably well and on similar teaching dimensions with evaluations by the deans and one of the peers. Recommendations for the construction and evaluation of teaching portfolios as well as for the use of student evaluations in summative evaluations are offered.

Paper presented
at
The American Educational Research Association Conference
April 1993

TM019116

The Use of the Teaching Portfolio and Student Evaluations for Summative Evaluation

John A. Centra

The teaching portfolio has been heralded as the latest contribution to effective teaching evaluation. Borrowed from such professions as art and architecture, in which professionals display examples of their work for prospective clients or employers, the concept is not totally new. Not long ago the same idea was called a teaching dossier, defined as a "summary of a professor's major teaching accomplishments and strengths" (Shore et al, 1986). Whether referred to as a portfolio, a dossier, or simply a faculty self-report, personal descriptions of teaching and other faculty activities should be the crux of summative evaluation. Most colleges, in fact, have for years included some type of teacher self-report or extended resume as a basis for personnel decisions. What is new are the kinds of information on teaching that are being promoted for inclusion in a "portfolio".

In the mid 1980s the Canadian Association of University Teachers sponsored a project to identify the kinds of information faculty members might use as evidence of their teaching effectiveness. Three major areas, which included 49 specific items, were included:

1. Material about students that reflects their learning (e.g., student workbooks or logs, student pre-and post-examination results)
2. Material from the faculty member (course materials, syllabi, descriptions of how various materials were used in teaching, innovations attempted and their evaluation, curriculum development)
3. Material from others (evaluations from students, colleagues or alumni), (Shore et al, 1986).

A portfolio could similarly include entries made by the professor alone or by others (Bird, 1990). Entries by a teacher could represent a wide range of practices, both good and bad, or the entries might be more selective and as some people have argued, display only the best work of a teacher (Wolf, 1991). In addition, most experts believe that a portfolio should include not only what teachers say about their teaching but what they actually do (Wolf, 1991; Edgerton, Hutchings and Quinlan, 1991). Moreover they argue that examples and artifacts should be included and that teachers' comments should emphasize why certain practices were followed. In this sense they are arguing that the portfolio should be

reflective and reveal what teachers were thinking and hoping for as they made instructional decisions. As Schon has discussed in The Reflective Practitioner (1983), professionals should not simply depend on established theory or technique but should react to particular situations that occur. Thinking and doing should not be separate; someone who reflects-in-action, Schon argues, becomes a researcher in the context of his or her job. The ideal portfolio would therefore highlight "a professor's reflections about a sample of actual work" Edgerton et al. (1991).

Lessons learned in portfolio design as part of the Stanford Teacher Assessment Project for K-12 teachers have been useful for college faculties as well (Bird, 1990; Wolf, 1991). Building on the Stanford project, Edgerton et al (1991) identified four domains that college professors could include in a portfolio. The first is course planning and preparation, represented by such work samples as course syllabi and lecture notes. The second is actual classroom instruction as represented, for example, by videotapes and colleague or student comments based on class observations. The third is evaluation and student feedback; the teacher's comments on a graded essay assignment is a sample of this third teacher task. The fourth domain is professional development in one's field--attending a professional conference, for example, and using the new knowledge gained in a course. For each of these domains, the teacher is expected to comment or reflect upon what was done.

Many colleges have in recent years used what they have defined as teaching portfolios for both formative and summative purposes. If they are used formatively, the information could facilitate self-analysis and improvement by capturing over time the teachers' descriptions of what they did in various courses and their reflections on their actions. Any judgements made by others would be offered as constructive suggestions. But if the portfolio is to be used summatively, judgements about what teachers have said and presented are not only necessary but may alter the contents of the portfolio.

Judgements of portfolio materials could be made by peers (as individuals, or as members of tenure/promotion committees), department chairs, deans, and other administrators. Because of the rich documentation that a portfolio can contain, the groups judging them would hopefully be in general agreement about the performance levels of individual teachers. These portfolio evaluations should also correlate with valid measures of teaching effectiveness, such as provided by student evaluations at the end of each course. Earlier studies of student evaluations have found significant correlations between selected items and measures of student learning in a course (Centra, 1977; Cohen, 1980). These results indicate that student evaluations reflect the amount learned in an instructor's section rather than, for example, the instructor's ability to entertain.

The purpose of this study was to investigate the possibilities and pitfalls of using portfolios for summative evaluations. A dean and two peers evaluated the portfolios prepared by faculty members at a college that required the portfolios for contract renewal purposes. Because student evaluations were also collected for each faculty member, this study was able to compare peer and dean judgements of teaching based on the portfolio contents with appropriate student evaluation scales and items.

Method

The college in this study was a community college that had used portfolios for two years and, during the second year, incorporated them into their faculty evaluation process. Each faculty member was asked to document his or her accomplishments and to write personal statements in four major areas: (1) teaching effectiveness, (2) service to the college and community, (3) personal credentials, and (4) professional activities. Teaching effectiveness was the most important category, receiving two-thirds of the weight in the compilation of a total score, while research and publications were excluded as a formal rating category. All together, raters could award up to 100 points for the categories in the self-report portfolio.

Teaching effectiveness was described by each faculty member under thirteen categories of performance grouped into three teaching skill areas: Motivational Skills, Interpersonal Skills, and Intellectual Skills. These skill areas and categories were adopted by the college from descriptions of teaching performance provided by Roeche and Baker (1987). A six point scale was used to rate each of the 13 teaching categories, ranging from "contradiction of the criterion (0)," and "criterion is not evident (1)," to "quality is strongly evident (5)." Thus, up to 65 points could be awarded for teaching by each rater.

Following is a list of the teaching skill categories with abbreviated examples of the kind of information teachers could provide for each:

Motivational Skills

1. Commitment to teaching: Availability to students, willingness to work on student clubs and activities
2. Goals orientation: Outlines goals and expectations for students
3. Integrated perception: Helps students link classroom experiences to the broader context of their lives
4. Positive action: Helps students achieve by motivating them with a desire to succeed
5. Reward orientation: Rewards received from teaching, signs of enthusiasm and satisfaction with teaching; how successful student performance is rewarded

Interpersonal Skills

6. Objectivity: Handles tough situations calmly and objectively, concentrating on the solution rather than the blame; uses communication skills effectively to involve students in the subject matter
7. Active listening: Paraphrasing for clarification, attending to non-verbal clues and demonstrating that what the student has to say is valued
8. Rapport: Achieving and maintaining a favorable relationship with students
9. Empathy: Reaching out to students in need and recognizing student feelings; expressing care yet asserting high expectations

Intellectual Skills

10. Individualized perception: Seeing students as individuals with different learning styles, different interests and different motivations, adjusting courses to individual needs
11. Teaching strategies: Employing a variety of well-organized teaching strategies; maintaining flexibility to be responsive to student needs

12. Knowledge: Staying current in your field and sharing the new knowledge with students in your classes; teaching from a wide range of sources including books, journals, conferences, etc.
13. Innovation: Integrating new ideas in a planned, deliberate way and willingly taking risks for a successful innovation (course syllabus should be attached)

Service to the college and community included activities for the past year only, with compensated responsibilities omitted. A point value of zero was allotted for no participation in either service area, while a point value of 15 was allotted for a continuous leadership role in two or more college service activities; five points were allotted for community service.

Personal credentials had a 10 point scale, with a doctorate or terminal degree in the teaching field receiving maximum value. Master's, bachelor's, associate's degrees and a certificate each received a decreasing number of points. A degree or certificate in a related teaching field was worth an additional point over one from an unrelated field.

Finally, professional activities had a maximum of 5 points for participation in professional organizations (inactive membership received no points, active participation in two or more organizations received four points, and leadership positions received five points).

Raters. Two peers and one of four deans rated each portfolio. One peer, herein designated Peer A, was selected by the individual faculty member as an appropriate judge; the second peer, designated Peer B, was selected by the area dean. The deans rated only the faculty members in their individual schools. In making their judgements, the raters relied heavily on the portfolios but did not have to limit themselves to what was written or included in them. After much discussion, the college faculty and staff decided that it would be difficult to exclude other perceptions or experiences they may have had with the person they were evaluating.

Student evaluations. The second source of information on teaching effectiveness was student evaluations collected at the end of a course. The college selected the Student Instructional Report, which is published by Educational Testing Service, for this purpose. Of the 39 items and six scale scores included in the SIR, two global items and three scale scores were emphasized by the college in the summative evaluations and were also especially appropriate for this study because they correlated reasonably well with student achievement in a previous study (Centra, 1977). The two global items, the overall value of the course and the overall quality of instruction, would be expected to correlate with the total Teaching Effectiveness score and the three teaching skill areas (Motivational Skills, Interpersonal Skills, and Intellectual Skills) in the portfolio. Three of the

SIR scales correspond to parts of the three teaching skill areas in the portfolio. The three SIR scales used for the evaluation were:

1. Organization and Planning: The extent to which teachers are perceived by students as well-organized; how well they prepare for each class, summarize major points in lectures or discussions and make their instructional objectives clear to students.

2. Faculty/Student Interaction: The extent to which instructors are perceived to be concerned with student progress and seem aware of when students need help; whether students also feel free to ask questions or to consult with the teacher.

3. Communication: Evaluations of the extent to which instructors raise challenging questions, use examples or illustrations, and give lectures of high quality.

The other three SIR scales excluded from the college's evaluations and also from this study were: Course Difficulty and Workload, Textbooks and Readings, and Tests and Exams.

Reliabilities for the SIR scales and items are good if the number of students making judgements is sufficient, as was the case for classes evaluated for this analysis. These reliabilities are reported elsewhere (Centra, 1973).

Sample. Virtually all full-time faculty at the college were evaluated during the 1990-91 academic year and were included in this analysis. They totalled 97 from four schools or divisions. In some cases just one class per faculty member was used for SIR ratings, but for the majority several classes were combined. The number of student ratings for each teacher ranged from 14 to 153, with an average of 52 students. Because of a change in governance of the college, including a name change, the evaluation information was to be used for contract renewal decisions for each faculty member. Thus, this presented a unique situation in which all faculty members were being summatively evaluated at the same time. In addition to evaluations of the portfolios by two peers and a dean, and the SIR results, each dean also made at least one unannounced visit to each teacher's classroom. These classroom visits were undoubtedly also taken into consideration in the deans' evaluations of portfolio information. The deans' evaluations were to receive 50 percent of the total weight, while peer and student evaluations each received 25 percent.

The data available from this college allowed a number of specific questions to be studied that shed light on the use of the teaching portfolio in summative decisions.

1. To what degree do ratings made by the two sets of peers and the deans differ?
2. How reliable are the ratings made by the peers and deans?
3. To what extent do the peer ratings agree (correlate) with each other and with the deans?
4. How do the evaluations of teaching made by the peers and deans (based largely but not entirely on self-reported information in the portfolios) compare with students' ratings on the SIR?

Results

The means and standard deviations for ratings given by the two sets of peers and the deans are given in Table 1. Ratings were made on the 13 various aspects of teaching effectiveness, college and community service, credentials, and professional associations. The mean ratings for all three groups of raters were uniformly high. For the two peer groups they ranged from 4.52 to 4.85 (out of a possible 5.00) on the teaching skill categories; for the deans they ranged from 3.75 to 4.49. The F-values, also listed in Table 1, indicate that the three groups of

raters did not differ significantly in their evaluation of each faculty member's credentials and the level of their participation in professional associations. They did, however, evaluate the dimensions of teaching as well as college and community service differently. At least two of the three groups of raters disagreed in their evaluations on these categories. The deans gave the lowest evaluations on each aspect of teaching and on the total teaching score (52.63 vs. 59.92 from Peer B and 62.00 from Peer A). The deans also rated College Service and Community Service lower than either set of peers. Of the two sets of peers, Peer A, selected by the faculty members, gave higher ratings than Peer B on total teaching as well as on the Motivation and Interpersonal Skills totals (indicated by the letter "b" next to the F-value). Thus the lowest ratings on teaching and service tended to be given by the deans, followed by the peer reviewers appointed by the deans.

Reliability of Ratings.

The reliability of ratings was estimated by the use of Coefficient Alpha, which measures the extent to which the individual categories in each of the teaching skill areas seem to be measuring the same concept. As indicated in Table 2, Coefficient Alphas were higher for the peers' than the deans' ratings. For Peer A and B ratings, they ranged from .70 on the Intellectual scale to .92 on the Teaching Total score. Most were

in the high .70s and .80s, suggesting that the categories within each scale were generally homogeneous and were rated with some consistency.

A few categories were especially influential in scale reliability as indicated by the size of Coefficient Alpha when a particular category is omitted. For example, omitting the Teacher Strategies category for Peer A ratings would reduce the Coefficient Alpha on the Intellectual Skills scale from .70 to .58. Thus the Teaching Strategies category was more influential than the other three categories included in the Intellectual than the other three categories included in the Intellectual scale for Peers A. This was also the case for the deans' ratings (from .37 to .12).

The deans' ratings had Coefficient Alphas of only .37 on the Intellectual Skills scale and .62 on the Motivational Skills scale. The Interpersonal and Teaching Total score reliabilities were higher at .70 and .79 respectively. For the deans, therefore, the Total Teaching score rather than the scales or individual categories provide the most reliable estimate of their evaluation of teaching effectiveness as described in the individual portfolios.

Intercorrelations Among Raters.

In Table 3 the correlations among the two peer groups and the deans are given for the three Teaching Skill scales and the

Total Teaching score. In general, Peer A did not correlate significantly with either the deans' evaluations or those of Peer B. The deans and Peer B evaluations did, however, correlate significantly with each other, with the correlation for Total Teaching highest at .43. Thus Peer B and the deans tended to be somewhat in agreement in their evaluations of each faculty member's teaching descriptions.

As shown in Table 4, the intercorrelations among the three groups of raters on College Service, Community Service, Personal Credentials, and Professional Activities were all significant ($p>.01$). Ratings of College Service correlated between .27 and .32, while ratings on the other three categories averaged about .50. Thus the peers and deans were in much greater agreement when they rated the more objective categories such as Personal Credentials, Community Service and Professional Activities.

Comparisons of Deans and Peer Evaluations of Portfolio-Reported Teaching Skills With Student Ratings on SIR

Means and standard deviations for the two global evaluation items on SIR and the three SIR scale scores are given in Table 5. These mean scores are at or just above the 50th percentile for a 1990 sample of two-year colleges and technical institutions. The SIR Interpretative Guide and Comparative Data provides mean scores based on responses from 86,816 students in 5,343 classes (see Educational Testing Service, 1990). The means of the 97

teachers in this sample on the "value of the course to students" was at the 60th percentile, while the rating of the quality of instruction was close to the median. One of the two scales was at the median while the other two were above the median. Instruction at this community college therefore, as rated by students, is generally in the mid-range relative to similar institutions.

Correlations between the deans' and peers' ratings on the teaching scales in the faculty portfolio with SIR items/scales are given in Table 6. Peer A ratings do not correlate significantly greater than zero with any of the SIR measures. Peer B and the deans' ratings on most of the teaching scales correlate significantly with three SIR measures: the quality of instruction item, the Faculty/Student Interaction Scale, and the Organization and Planning Scale. Neither the SIR Communications scale nor the SIR item rating the value of the course had consistently high correlations with portfolio categories, except for the Motivational Skills Scale; both Peer B and the deans' ratings on this category correlated with the student ratings of the value of the course. These correlations tell us something about the three sets of ratings and the content of the scales, which is discussed below.

Discussion

The results of this study have significance for the construction and use of faculty portfolios, particularly the descriptions and reflections on teaching, which are a key aspect of a portfolio. This study also sheds additional light on the validity of student, peer, and administrator evaluations of teaching.

The faculty portfolio used by the college in this study included descriptive, evaluative, and reflective information provided by each faculty member. Most of the information dealt with teaching, but participation in service to the college and community, and in professional associations was also included. Because the results were to be used for summative decisions on each member of the faculty, great care was generally taken in preparing the portfolio. The faculty provided specific examples and descriptions of commitment to teaching, their involvement with students in the subject matter, their willingness to be flexible in response to student needs, and other categories that reflected teaching skills. Only positive examples were requested, so it is not surprising that peers and deans rated performance highly overall: on a six point scale most ratings were at or above four. When a portfolio is being used for summative decisions, it is reasonable to ask teachers to provide only positive examples of their effectiveness. For formative

purposes, however, reflections on how one may have done better would be less threatening to an individual and could be useful in improvement.

Rater Effects

For summative purposes, evaluation of the contents of a portfolio are critical to the personnel decisions being made. Who makes those evaluations is also critical, as this study indicated. Peers selected by the individual faculty member were the most lenient in their evaluations. They differed from the dean and to some extent from the peer chosen by the dean. Most likely the fact that each peer was also to be evaluated, caused peers to be less critical. An earlier study of peer evaluations in which peers judged each other also produced very high peer evaluations (Centra, 1975). When peers are on a tenure and promotion committee (or an ad hoc committee to evaluate a candidate's teaching), and are not being evaluated simultaneously, they might be expected to be somewhat more objective in their evaluations.

The peers chosen by the dean would presumably be less influenced by personal associations with the teacher being evaluated or by other biasing factors. Perhaps a more random selection of these peers would also ensure that they would not be influenced by the views of the dean.

None of the groups of raters differed in their evaluations of credentials and participation in professional associations, indicating that systematic bias or differing points of view occurred only in evaluating teaching and service. Not only did each group assign similar mean values, but the intercorrelation among the groups for Credentials and Professional Associations were fairly high (.35 to .68, in Table 4). The intercorrelations among the groups were also significant for the two service areas (.27 to .66, Table 4). This indicates that the relative judgements made by the three groups of raters were fairly similar. Even though the peer groups, particularly those selected by the faculty members, gave higher ratings, there was a significant similarity in how they ranked faculty members in these four areas of the portfolio: credentials, professional activities, community service, and to a lesser extent, college service.

For teaching, however, only the deans and the peers selected by the deans (Peer B) gave similar relative judgements (Table 3). The opinions of the peers named by the faculty member being evaluated (Peer A) differed from the others, suggesting that these were the least valid evaluations. This invalidity, or lack of agreement with others, was evident in the student evaluation results as well: Peer A evaluations did not correlate with any of the SIR scales or items.

SIR Evaluations

The SIR scales and items that correlated most consistently with Peer B and dean evaluations of teaching were the Organization and Planning scale, the Faculty/Student Interaction scale, and the Overall Quality of Instruction item. These three parts of SIR would therefore appear to come closest to measuring the three teaching skill areas reflected in the teaching portfolio.

The SIR item that rated the value of the course to students did correlate with the Motivation Skills area of the portfolio, which included the extent to which instructors help students link classroom experiences to life. The Communications scale, however, correlated only modestly with the deans rating on Intellectual Skills, which included the extent to which a variety of teaching strategies are used and the extent that new knowledge is shared effectively with students.

The SIR Organization and Planning scale reflects students' views of a well-organized, well-prepared teacher who makes course content clear by giving examples and specifying objectives. The portfolio category of Motivational Skills reflects similar ideas: outlining course goals, motivating students to succeed, and rewarding students for successful performance. The correlation between the SIR Planning Scale and this portfolio category reflects this agreement.

The SIR Faculty/Student Interaction scale, with its emphasis on concern for students, overlaps all three of the teaching skill areas in the portfolio: Motivational, Interpersonal, and Intellectual. Thus, the Faculty/Student Interaction scale would be expected to correlate with these teaching skills evaluated in the portfolio, as indeed it did. In sum, the SIR student evaluations correlated reasonably well and on similar teaching dimensions evaluated by deans and peers, i.e. Peer B. Most previous studies that compared student, peer, and administrator evaluations used only global or overall evaluations of teaching. In these studies, peers and administrators based their ratings on reputations, hearsay, or other unspecified sources of evidence. Only student evaluations were based on classroom performance. In his review of 14 of these studies, Feldman (1989) reported an average of .55 between peer and student evaluations. Peer and administrator (deans and department chairs) evaluations of teachers correlated .48 (five studies). Administrator ratings of teachers correlated .39 with those by students (11 studies). These correlations are slightly higher than those found in this study. Basing evaluations on a portfolio, particularly for summative purposes, apparently introduces other sources of error. For example, the peers and deans were largely expected to use their own criteria and standards for judging the portfolios. Moreover, the portfolio required by the college and used in this study did not include many work samples that could represent

teaching performance. Examples provided by Edgerton et al (1991) and Seldin (1991) include such items as:

-a personal statement by the teacher describing instructional goals for the next several years
-representative course syllabi (requested of faculty in this study)
-examples of graded student essays
-hard evidence of student learning (examination scores pre-and post course)
-a videotape of the professor teaching a course

The portfolios tended to focus on the responses given by each faculty member in 13 teaching categories identified by Roeche and Baker (1987) and modified by the college. The 13 categories were grouped into three skill areas--Motivation, Interpersonal, and Intellectual. The Coefficient Alphas were acceptable for the two peer groups but not the deans, indicating that for these administrators the teaching skills categories did not generally fall within the skill area designated (in particular for Intellectual and Motivation skills). Individualized Perception (seeing students as sharing different learning styles and motivations), for example, could just as easily be a Motivational Skill as an Intellectual Skill. Overlap between such categories as Rapport and Empathy also frustrated

teachers or caused them to repeat themselves. Fewer and more sharply distinguished categories for each skill area would be easier for both teachers and evaluators.

The evaluations of the portfolios in this study would have undoubtedly benefitted from additional discussion among the evaluators about the criteria and standards to apply. In a study in which six elected members of the faculty rated faculty dossiers after first discussing the criteria and examples of high and low ratings, the agreement among the peers was very high (Root, 1987). The dossiers included various teaching materials as well as student evaluations. They also included publications and grant proposals used to evaluate research, and documentation of service activities. In the Root study reliabilities of peer evaluations in all three areas was above .90, indicating that the peers gave essentially the same ratings to the faculty they rated. The strongest agreement was in evaluating research, the lowest in service. The brief "training" that took place undoubtedly contributed to the higher rating correlations between peers in the Root study than in the study reported here. Future evaluation of portfolios or dossiers by peers or various administrators should consider including written criteria or group discussions about the common standards and criteria to apply. Doing so would no doubt lead to greater agreement among raters.

The ideal portfolio is put together by a teacher over a period of several years. Because of the college's need to make immediate use of portfolios as part of a total faculty evaluation process, the faculty in this study did not have the opportunity to do so. Thus the portfolio was more like a snapshot of teaching performance, albeit with much descriptive detail, than a longitudinal, documented set of changes or results over time.

Nevertheless, when summative decisions are being made, even portfolio procedures that are not ideally designed can assist evaluators of teaching performance. And when those evaluations are combined with valid assessments of teaching by students, a multiple perspective on teaching effectiveness is provided.

References

Bird, T. "The Schoolteacher's Portfolio: An Essay On Possibilities" in The New Handbook of Teacher Evaluation: Assessing Elementary and Secondary School Teachers. Ed. by J. Millman and L. Darling-Hammond, 241-256, Newbury Park, Calif: Sage, 1990.

Centra, J. A. The Student Instructional Report Number 3. Princeton, N.J.: Educational Testing Service, 1973.

Centra, J.A. "Colleagues as Raters of Classroom Instruction." Journal of Higher Education, 1975, 46, 327-337.

Centra, J. A. "Student Ratings of Instruction and Their Relationship to Student Learning." American Educational Research Journal, 1977, 14(1), 17-24.

Cohen, P. A. "Effectiveness of Student Rating Feedback for Improving College Instruction: A Meta-Analysis of Findings." Research in Higher Education, 1980, 13, 321-341.

Edgerton, R., Hutchings, P. and Quinlan, K. The Teaching Portfolio: Capturing The Scholarship in Teaching. Washington, D.C.: The American Association for Higher Education, 1991.

Educational Testing Service, The SIR Interpretative Guide and Comparative Data. Princeton, N.J.: ETS, 1990.

Feldman, K. A. "Instructional Effectiveness of College Teachers as Judged by Teachers Themselves, Current and Former Students, Colleagues, Administrators and External (Neutral) Observers", Research in Higher Education, 1989, 30.

Roeche, J. and Baker, G. Access and Excellence: The Open Door College. Washington, D.C.: The Community College Press, 1987.

Root, L. S. "Faculty Evaluation: Reliability of Peer Assessments of Research, Training, and Service," Research in Higher Education, 1987, 26, 71-84.

Schon, D. The Reflective Practitioner, New York: Basic Books, 1983.

Seldin, P. The Teaching Portfolio. Boston, MA: Anker Publishing Co., Inc. 1991.

Shore et al, The Teaching Dossier: A Guide to Its Preparation and Use. Revised Edition. Montreal: Canadian Association of University Teachers, 1986.

Wolf, K. "The Schoolteacher's Portfolio: Practical Issues in Design, Implementation, and Evaluation," Phi Delta Kappa. October, 1991, 129-136.

Zuckerman, H. A. and Merton, R. K. "Age, Ageing and Age Structure in Science." In R. K. Merton, The Sociology of Science: Theoretical Empirical Investigations, edited by N. Storer. Chicago: University of Chicago Press, 1973, 497-559.

Zuckerman, H. "The Sociology of Science." In Handbook of Sociology, edited by N. J. Smelser, Newbury Park, Calif: Sage, 1988, 511-574.

TABLE 1

Means, Standard Deviations and F-Values Among Raters on Specific Variables

N=97

	Peer A N=97		Peer B N=97		Deans N=4		
	Mean	S.D	Mean	S.D.	Mean	S.D.	F-Value
1Commitment	4.85	.54	4.72	.67	4.31	1.03	11.09a
2Goal Orientation	4.83	.53	4.58	.88	4.01	1.03	24.40b
3Integrated Perception	4.77	.59	4.63	.81	4.08	1.01	17.80a
4Positive Action	4.74	.70	4.56	.80	4.00	1.16	16.69a
5Reward Orientation	4.77	.60	4.61	.76	3.75	1.02	30.23a
6Objectivity	4.73	.67	4.51	.89	4.23	.99	7.83a
7Active Listening	4.75	.65	4.58	.85	4.07	1.03	16.65a
8Rapport	4.84	.55	4.63	.78	4.03	1.06	22.72b
9Empathy	4.81	.55	4.68	.70	3.89	1.15	24.74a
10Individualized Percep.	4.77	.59	4.56	.85	4.02	1.02	23.30b
11Teaching Strategies	4.71	.68	4.54	.87	3.78	1.02	28.48a
12Knowledge	4.77	.60	4.79	.59	4.49	.89	4.67a
13Innovation	4.58	.83	4.52	.91	3.96	1.04	12.64a
14College Service	12.65	3.32	12.56	3.17	10.32	3.53	20.73a
15Community Service	4.14	1.15	4.15	1.20	3.63	1.47	7.91a
16Credentials	8.62	1.73	8.69	1.52	8.64	1.30	.39
17Professional Assoc.	4.06	1.16	4.00	1.16	3.77	1.43	2.18
Total Score	91.69	8.59	89.32	10.68	78.94	10.52	72.88b
18Motivation Total	23.99	2.36	22.98	3.08	20.11	3.33	48.99b
19Interpersonal Total	19.15	1.85	18.54	2.63	16.19	3.11	34.86b
20Intellectual Total	18.86	1.94	18.40	2.54	16.23	2.32	49.26a
Total Teaching	62.00	5.59	59.92	7.42	52.63	7.16	65.04b

a Overall F significant at .05 level. Post hoc comparisons using MANOVA identified significant differences between either of the peers and the dean.

b Overall F significant at .05 level. Post hoc comparisons MANOVA identified significant differences between all pairs of raters

TABLE 2

Coefficient Alphas Among Raters for Motivation, Interpersonal, Intellectual, and Total Teaching

N=97

	Peer A	Peer B	Deans
Motivational Skills (Coefficient alpha when identified variable is omitted.)	.85	.77	.62
1Commitment	.82	.74	.51
2Goal Orientation	.83	.72	.62
3Integrated Perception	.83	.72	.57
4Positive Action	.80	.69	.50
5Reward Orientation	.83	.76	.59
Interpersonal Skills (Coefficient alpha when identified variable is omitted.)	.77	.87	.70
6Objectivity	.65	.85	.62
7Active Listening	.75	.83	.58
8Rapport	.65	.85	.67
9Empathy	.78	.83	.68
Intellectual Skills (Coefficient alpha when identified variable is omitted.)	.70	.79	.37
10Individualized Per.	.62	.71	.40
11Teaching Strategies	.58	.73	.12
12Knowledge	.67	.78	.31
13Innovation	.68	.72	.36
Teaching Total	.91	.92	.79

TABLE 3

Correlations Among Raters for
Motivational Skills, Interpersonal Skills, Intellectual Skills, and Total Teaching
N=97

	Deans				Peer A			
	Motivation	Interpersonal	Intellectual	Total Teaching	Motivation	Interpersonal	Intellectual	Total Teaching
Peer A								
Motivation	.00							
Interpersonal		.04						
Intellectual			.22*					
Total Teaching				.04				
Peer B								
Motivation	.40**				.14			
Interpersonal		.39**				.16		
Intellectual			.24*				.19	
Total Teaching				.43**				.17

* $p < .05$
** $p < .01$

TABLE 4

Correlations Among Raters for Scores in
College Service, Community Service, Credentials, and
Professional Activities
N=97

	Deans				Peer A			
	(I) College	(II) Community	(III) Credentials	(IV) Professional	(I) College	(II) Community	(III) Credentials	(IV) Professional
Peer A								
I	.32**							
II		.47**						
III			.65**					
IV				.44**				
Peer B								
I	.29**				.27**			
II		.52**				.66**		
III			.68**				.55**	
IV				.40**				.35**

* $p < .05$
** $p < .01$

TABLE 5

Means and Standard Deviations on SIR Items/Scales

N=97

	Mean[1]	Percentile[2]	Standard Deviation
SIR Overall Value of Course to Students (Item)	4.24	60	.32
SIR Overall Quality of Instruction to Students (Item)	4.24	48	.34
SIR Communication Scale	9.86	50	.52
SIR Planning Scale	10.54	53	.64
SIR Interaction Scale	10.72	63	.72

[1]Mean of ratings for 97 teachers, whose ratings were based on between 14 and 153 students in one or more classes.

[2] Based on 1990 SIR Comparative Data for Two Year Colleges (p. 45, Educational Testing Service, 1990), and on scale score distributions provided by ETS.

TABLE 6

Correlations Between
Motivational Skills, Interpersonal Skills, Intellectual Skills, Total Teaching and SIR Items/Scales

N=97

	SIR Value of Course	SIR Quality of Instruction	SIR Communication Scale	SIR Interaction Scale	SIR Planing Scale
Peer A					
Motivation	.00	.02	-.14	-.06	-.07
Interpersonal	-.06	-.11	-.20	-.11	-.13
Intellectual	.07	.05	.08	.04	.00
Total Teaching	.00	-.03	-.10	-.04	-.07
Peer B					
Motivation	.20*	.33**	.03	.34**	.33**
Interpersonal	.14	.27**	.07	.31**	.29**
Intellectual	.13	.27**	.09	.28**	.25*
Total Teaching	.17	.33**	.07	.35**	.33**
Deans					
Motivation	.28**	.34**	.11	.38**	.35**
Interpersonal	.10	.19	.03	.24*	.18
Intellectual	.18	.20	.21*	.29**	.20
Total Teaching	.26*	.33**	.15	.40**	.33**

* $p < .05$
** $p < .01$